

Threats to Reliability and Validity

EDU 500

Compensation

Things weren't equal with my groups

I felt so bad I could cry

I tried to balance for the unlucky ones

And watched my validity fly (away)



Example:

Hold review/tutoring sessions for one group and withhold from another. Feel bad and give group without sessions a study packet.

Setting

Drug/Alcohol Use? Cheating? Sex & Sexuality?

A study promising confidentiality

Face-to-face survey, Others walk past

Truthful answers disappear fast



Maturation is present when a physical or mental change occurs over time and it affects the participants' performance on the dependent variable.

Educational Example:

If you measure first grade students' ability to perform arithmetic problems at the beginning of the year and again at the end of the year, some of their improvement will probably be due to their natural maturation (and not just due to what you have taught them during the year). Therefore in the one group design, you will not know if their improvement is due to the teacher or if it is due to maturation.

Maturation is not a threat in the two group design because as long as the people in both groups mature at the same rate, the difference between the two groups will not be due to maturation.

“You may grow, you may change,
but your answers will probably not be the same.”



Apprehension: dread, a feeling of anxiety that something bad or unpleasant is going to happen.

“Are your eyes twitching, do you have an upset stomach, or a headache?” “Then, you might want some tums, Mylanta or Peptobismal?”



Educational relevance: Test-taking anxieties—timed or state tests. The higher performing students sometimes score significantly lower compared to their everyday tasks and practices.

History

History refers to the occurrence of events that could alter the outcome or the results of the study.

Example- Due to the budget crisis of 2008, many schools cut back resources.



Instrumentation

Catch Phrase:

Switching up is not a way to be.
Keep the same and stay with Validity.

Example:

The person giving the test changes from a woman to a man.

Statistical Regression

Less is more, but NOT when it pertains to average scores!



Differential Selection of Partners

Differential selection is defined as selecting participants for various treatment groups that have different characteristics.

When you have two or more groups (and no random assignment to the groups), any difference observed between the groups might be due to the characteristics of the people in the different groups rather than the treatment. In other words, the selection variables might be the real reason that the groups differ. In short, you cannot conclude that the observed differences between the groups at the posttest is due to the different treatments because it is confounded with participant characteristics

Ex: Pre-formed groups = Honors Class vs. Academic Class

- Results on post-test might be a result of the characteristics of the pre-formed groups

Ambiguity

The glass is half empty or maybe half full.
Can I tell the difference with a measuring tool?
Both are correct it seems to me.
A possible threat to validity?

Rubrics can often
contain ambiguity.
Ex. Student “effectively”
copes with change.

OVERBOARD



Definition: Improvement due to an expectation rather than treatment itself; can occur when participant receive a treatment that they believe to be beneficial

Catch Phrase:

It may not be happening to you,

But you will believe it is true.

The improvements you' ll see from a placebo are uncanny,

But the actual results you' ll have...there aren' t any.

Placebo Effect



Educational Example:

You have 4 groups of 6th graders, two groups of experimental and two groups of control, and the treatment is a program to promote a healthy lifestyle, such as “DARE” or “Just Say No.” The experimental groups are excused from several classes to view the program. The control groups are also excused from several classes to view a different program which is unrelated to the study. As an additional control, you might tell all students that there are two programs and that eventually they will all see both programs. This way it will ensure that all students are doing the same thing. All students were polled at a later point in time and reported living a healthy lifestyle.

Selection-Maturation Interaction

Selection maturation is a threat

It will affect your results, I bet.

Let's say you choose young kids for your test

Time and growth may change their reaction at best



Example: A group of children practices doing the monkey bars over a period of 2 years. Do some of the children get better from practicing or because they grew taller than the other children?

Multiple Treatment Interference

When two or more treatments are repeated, the results can not be generalized to a single treatment, so the validity may be defeated. Carryover effects from an earlier treatment may make it difficult to assess the effectiveness of a secondary treatment.



Educational example:

Comparing 2 different approaches to improving classroom behavior.

Behavior modification the first 2 months and corporal punishment the second 2 months. Behavior improved with both, but it is difficult to determine if the improved behavior was a carryover from the behavior modification treatment, or corporal punishment treatment.



Experimenter Bias

The Outcome is Suggested by the Experimenter

Subjective not Objective because it

Uses subtle communication to

Guide the results

Getting invalid data because

Experimenter has a preference

Study can avoid this bias by

Taking advantage of blind study method

Educational Example: An experimenter's review of a short answer response, of students they know, may be influenced by what they know about the student.

Hawthorne Effect

**Show me
some
attention
...**



**AND....I
will
perform
for you!**



A term referring to the tendency of some people to work harder and perform better when they are participants in an experiment. Individuals may change their behavior due to the attention they are receiving from researchers rather than because of any manipulation of independent variables.

Educational Example: Give extra attention to that child who needs to perform better academically and see if he/she performs better.

Novelty effect, in the context of human performance, is the tendency for performance to initially improve, increased interest, motivation and engagement when doing something new or different, not because of any actual improvement in learning or achievement.

HAVE THE BEST BRIDE TO BE PARTY.
INVITE ALL OF YOUR FRIENDS AND RACY AUNTS.
MONEY WELL SPENT.....
DECORATIONS, PARTY GAMES, GIFTS FOR THIS ONE TIME EVENT.



Educational example: The increased attention by students sometimes results in increased effort or persistence, which yields achievement gains. If they are due to a **novelty effect**, these gains tend to diminish as students become more familiar with the new medium.

To counteract the effect, study should be conducted over a period of time sufficient to allow “newness” to wear off.

John Henry Effect

Catch Phrase:

I' m in a control group, that I **do** know-
So **changing** my behavior is the way that I go...
For I want certain results, to **benefit** me-
To the heck with the researchers' **validi-tee**

Educational Example:

An IEP student who knows that they are part of a control group is assigned a one-on-one aide. This assignment is meant as a short term fix to improve performance in a specific class, specific unit. They purposely misbehave and perform poorly in an effort to keep the assigned aide for added support for a longer time period.

Interaction



Did I kill 2 birds with 1 stone? OR did
1 die on its own?

Mortality/Attrition

We refuse! We want to move!

If we bail the validity will fail!

Then the data you can no longer use!

We retorted, now the research is distorted!



Attrition occurs when participants drop out of the study before all data is collected, thus altering the accuracy of results. Subject mortality can stem from a participant refusing participation because of the sensitive nature of the data being collected; participants may relocate between data collection times; data collection procedures may be costly to a participant; or schedules may change. The most important concern of mortality is that the loss of subjects possibly undermines the credibility of results.

Educational Example:

A school wants to examine the success of their after-school remediation program. To do so they want to administer and compare pre-test and post-test results. However, some of the students stop participating in the program a few weeks in. Now the average gains from pretest to posttest will be invalid.

Diffusion – when effects spread between the experimental and control groups lessening the difference between them because the control group has figured out the experiment and started mimicking the symptoms or the dependent variable

Catch Phrase:

I didn't feel a thing before we talked.

But I think I understand what you said.

Now my stomach hurts and I'm sneezing.

No, no, no. It isn't in my head.

Visual:



Relevant educational example: A teacher could decide to test the success of two different ways to study for vocab tests. One group is given flash cards and told to study on their own using the cards. A second group is put into pairs and allowed to quiz each other in order to prepare. Diffusion occurs when the flash card kids begin asking to study in pairs because they deem the group work more desirable.

Rivalry

Definition: The comparison group knows what the program group is getting and develops a competitive attitude what the program group

Educational Example: The students in the comparison group might see the special math-tutoring program the other group is getting and feel jealous. This could lead them to compete with the program group "just to show" how well they can do.

“Competition offers two possibilities.
You could lose.
Or, if you want to win, you can change.”



Testing/Pre-testing Sensitization

Testing, testing, testing, when we pre-test our subjects can be sensitized to many pressures. Teachers, peers, subject matter, general surroundings causing crazy results on the real test or the post test!



Selection



If you don't use a random sample, you won't get valid results.

But, even random samples can have their faults,
Too many girls, not enough boys
all of the external variables will act as decoys.

Definition: The validity of research can be threatened based on the sample. If a nonrandom group of people take part in the study they may have similarities that impact the outcome of the research. Even in random samples, an imbalance of diversity will often occur (ex. Out of 100 people randomly selected there are a lot more females than males)

Educational Example:

High school graduates are randomly selected to complete a "post-graduation" survey about their high school experience weeks after graduation. Despite the randomness of the subjects selected, less than half of the surveys were returned. The majority of these surveys reported a positive high school experience.

What personality traits would a person possess that would lead them to complete this survey? Could these personality traits also impact the high school experience of these students?